ELSEVIER

Biochimica et Biophysica Acta 1338 (1997) 161–185

BIOCHIMICA ET BIOPHYSICA ACTA
BBA

Review

The application of circular dichroism to studies of protein folding and unfolding

Sharon M. Kelly, Nicholas C. Price *

Department of Biological and Molecular Sciences, University of Stirling, Stirling FK9 4LA, Scotland, UK

Received 18 June 1996; revised 16 October 1996; accepted 16 October 1996

Contents

Abbreviations: CD, circular dichroism; GdnHCl, guanidinium chloride; MRW, mean residue weight; ANS, 1-anilino-8-naphthalene-sulfonate; Mb, myoglobin; FTIR, Fourier transform infra red; NOE, nuclear Overhauser enhancement; BPTI, bovine pancreatic trypsin inhibitor; LDAO, lauryldimethylamine-*N*-oxide

* Corresponding author. Fax: +44 1786 464994; E-mail: ncp1@stirling.ac.uk

PII S0167-4838(96)00190-2

1. Introduction

In this review, we shall outline the basic principles of circular dichroism (CD) and indicate the types of structural information relevant to the study of the folding and unfolding (or denaturation) of proteins which can be obtained. We shall then describe a number of studies of these processes, using systems ranging from small peptides to complex multi-domain and multi-subunit proteins, which will indicate the particular contribution which CD can make in conjunction with other structural techniques. Finally, the way in which this information has been used to help formulate models for protein folding will be outlined.

1.1. Basic principles of CD

CD refers to the differential absorption of the left and right circularly polarised components of plane-polarised radiation. This effect will occur when a chromophore is chiral (optically active) either (a) intrinsically by reason of its structure or (b) by being covalently linked to a chiral centre or (c) by being placed in an asymmetric environment. In practice radiation is split into the two circularly polarised components by passage through a *modulator* (usually a piezoelectric crystal such as quartz) subjected to an alternating (50 kHz) electric field. The modulator will transmit each of the two components in turn. If, after passage through the sample, the components are not absorbed (or are absorbed to the same extent), combination of the components would regenerate radiation polarised in the original plane. However, if one of the components is absorbed by the sample to a greater extent than the other, the resultant (combined component) radiation would now be *elliptically polarised*, i.e., the resultant would trace out an ellipse (Fig. 1). In practice, the CD instrument (spectropolarimeter) does not recombine the components but detects the two components separately; it will then display the dichroism at a given wavelength of radiation expressed as either the *difference in absorbance* of the two components ($\Delta A = A_L - A_R$) or as the *ellipticity* in degrees (θ) ($\theta = \tan^{-1}(b/a)$, where b and a are the minor and major axes of the resultant ellipse (Fig. 1). There is a simple numerical relationship between ΔA and θ (θ in degrees), i.e., $\theta = 32.98\,\Delta A$.

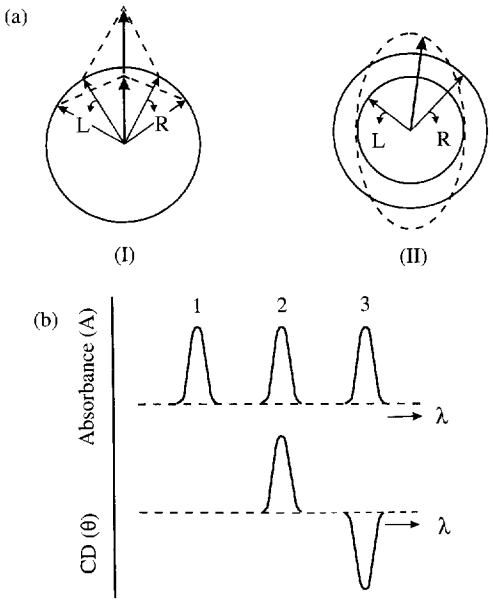

Fig. 1. Origin of the CD effect. (a) The left (L) and right (R) circularly polarised components of plane polarised radiation: (I) the two components have the same amplitude and when combined generate plane polarised radiation; (II) the components are of different magnitude and the resultant (dashed line) is elliptically polarised. (b) The relationship between absorption and CD spectra. Band 1 is not chiral; band 2 has a positive CD spectrum with L absorbed more than R; band 3 has a negative CD spectrum.

In practice in most biological work, the observed ellipticities are of the order of 10 millidegrees (i.e., the difference in absorbance between the two circularly polarised components of the incident radiation is of the order of 3×10^{-4} absorbance units). It is therefore clear that careful attention must be paid to the experimental conditions to ensure that meaningful data are obtained (Section 2.2). A CD spectrum is obtained when the dichroism is measured as a function of wavelength (Fig. 1). As discussed in more detail in Section 2.1, the CD spectra of proteins can give insights into a number of aspects of protein structure, and when used in combination with other structural techniques can play an important role in examining the processes of the folding and unfolding of proteins. A number of excellent reviews of CD have been published (e.g., [1,2]) and recently a comprehensive multi-author treatise has appeared [3].

1.2. The folding and unfolding of proteins

The biological properties of a protein depend on its three dimensional structure and the way in which that structure can be influenced by chemical and physical factors (ligands, pH, temperature, etc.). Although the structures of over a thousand proteins are now known as a result of the application of techniques such as X-ray crystallography or high resolution NMR, much less is known about the pathway of protein folding. It is clear that in vivo such folding (and assembly in the case of multi-subunit proteins) takes place both *rapidly* (on time scale of seconds) and *efficiently* (i.e., with little or no by-products such as aggregated protein). The most useful approach to examining the process of protein folding has been to study the refolding of proteins after unfolding; the extent to which this in vitro approach is a suitable model for protein folding in vivo has been discussed at length [4,5]. It is clear that at least for a number of small proteins both the rate and efficiency of refolding are compatible with those presumed to occur in vivo. For larger multi-domain or multi-subunit proteins the rate of refolding can be much slower and the efficiency of the process relatively low, with formation of aggregates a common competing process. This suggests that, in vivo, there are various factors which can improve the folding process in terms of rate and efficiency. Since about the mid-1980s the role of a number of such cellular protein factors, generally termed collectively 'molecular chaperones', has been elucidated (for reviews see [6–8]). It has become clear that the folding of any given protein, especially one which is translocated from its site of synthesis, involves the sequential action of a number of different chaperones.

Deciphering the mechanism of protein folding and hence establishing the detailed relationships between sequence and three dimensional structure is not only of considerable intrinsic biological interest (the process has been termed 'the second half of the genetic code'), but is also of considerable technological and clinical importance. In many cases, the overexpression of recombinant proteins leads to the formation of 'inclusion bodies' containing the desired protein in an aggregated insoluble form [9]. Recovery of the protein in an active state involves solubilisation in a denaturing agent (usually urea or guanidinium chloride (GdnHCl)) followed by refolding under carefully controlled conditions; this can often be the most difficult step in the entire production process [10]. In addition refinements in the chemical methods of peptide synthesis have allowed 'designed' proteins of moderate size (up to about 100 residues) to be produced in reasonable yield [11]; this approach offers the advantage that unnatural amino acids such as norvaline can be incorporated into proteins. There is thus a pressing need: (a) to understand the process of protein folding, (b) to be able to predict the three dimensional structures of proteins from their amino acid sequences, and (c) to be able to calculate the effects of changes in the amino acid sequence on the structure, and hence on the properties of the protein. The clinical relevance of understanding protein folding pathways is well illustrated by the fact that a number of disease states (e.g., cystic fibrosis, scurvy, maple syrup urine disease, α_1-antitrypsin deficiency) may be linked to the formation of misfolded proteins, which can then lead to incorrect assembly or mislocalisation [12,13].

As a result of extensive studies, mainly on the refolding of small proteins, a number of important insights into the folding pathways of proteins have been obtained, with the delineation of some structural features of intermediates which may be involved and the measurement of the rates of key processes. CD has played an important role in much of this work, as outlined in Sections 4.3 and 4.4.

The native three-dimensional structures of proteins are of only relatively limited stability (typical values are 20–60 kJ mol^{-1}) [14]; this is a consequence of the relatively close balance between those forces (hydrophobic, hydrogen bonding, ionic bonding and van der Waals' forces) tending to promote a folded structure and the entropic factors which favour the unfolded state. Detailed measurements of the unfolding of proteins are important in order to define the range of conditions under which native-like biological properties are retained. Measurements of the structural changes in proteins by denaturants such as urea or GdnHCl can be used to establish the validity of models for unfolding, and in favourable cases can give quantitative estimates of the free energy difference $\Delta G°(H_2O)$ between the folded and unfolded states. These estimates could be used for instance to give a direct assessment of the effect of an amino

acid substitution on the stability of a protein (Section 3.3), and hence an insight into the role of that amino acid in crucial interactions which stabilise the folded state. Finally, studies of the unfolding of proteins and peptides may also reveal information on the properties of partially folded intermediate states, which could be relevant in understanding the pathway of folding (Sections 3.4, 3.5 and 3.6).

2. Structural information and spectroscopy

2.1. Structural information on proteins available from CD

One of the strengths of CD is that various aspects of protein structure can be measured. Studies of the far-UV region (typically 240 nm to 190 or 180 nm) can be used to assess quantitatively the overall secondary structure content of the protein. In this region, the absorbing group is principally the peptide bond. There is a weak but broad $n \rightarrow \pi^*$ transition centred around 210 nm and an intense $\pi \rightarrow \pi^*$ transition about 190 nm. In the near-UV, the aromatic amino acid side chains (phenylalanine, tyrosine and tryptophan) absorb in the range 250 to 290 nm; the tertiary folding of the polypeptide chain can place these side chains in chiral environments, thus giving rise to CD spectra which can serve as characteristic fingerprints of the native structure. In model compounds containing a disulfide bond, the dihedral angle of this bond can often be deduced from the appearance of the CD spectrum in the 240 to 290 nm region. In the case of extensively disulfide bonded proteins, there can be difficulties in performing such analyses because of the overlap of the disulfide signals with others from the aromatic side chains [1,2,15–17]. Finally, non-protein components or cofactors, such as flavins, haem, pyridoxal 5′-phosphate, etc., may absorb in regions of the spectrum well separated from those of amino acids and peptide bonds; the CD signals in these regions can be used to provide detailed information on the environment of these chromophores. Each of these aspects will now be discussed in turn.

2.1.1. The far-UV CD of proteins (peptide bond dichroism)

It has been known for many years that the different forms of regular secondary structure found in peptides and proteins exhibit distinct far-UV CD spectra (Fig. 2). The task of taking any observed spectrum and deducing the contributions of the different structural forms would thus appear to be limited to the solution of a (large) number of simultaneous equations using data over the far-UV. In practice, this has turned out to be more difficult for the following reasons. Firstly, it is clear that aromatic amino acid side chains (especially tryptophan) can contribute to the far-UV spectrum (Section 2.3) [2,18,19]. Secondly, disulfide bonds can contribute to the spectrum [1,2,15–17]. Thirdly, the characteristics of the spectra can depend on the length and regularity of structural elements in peptides and proteins [20]. More recent approaches have been based on sophisticated curve-fitting procedures involving sets of spectra of well-characterised proteins of known secondary structure. The CONTIN procedure introduced by Provencher and Glöckner [21] involves the direct analysis of a CD spectrum over the range from 240 to 190 nm as a linear combination of the CD spectra of 16 proteins whose structures have been determined to high resolution by X-ray crystallography. The approach avoids

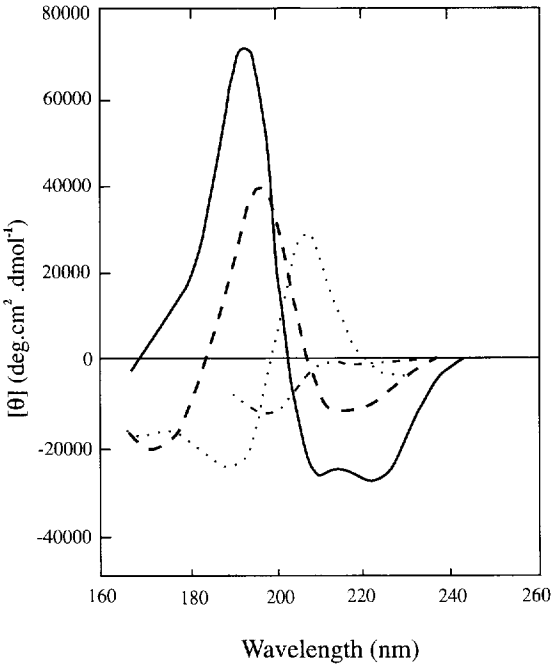

Fig. 2. Far-UV CD spectra associated with various types of secondary structure: solid curve, α-helix; long dashes, anti-parallel β-sheet; dots, type I β-turn; dots and short dashes, irregular structure [23].

the problem of defining reference spectra and provides a flexible analysis of actual protein spectra. In the CONTIN procedure, greater weight is automatically given to those proteins in the reference set whose spectra more closely resemble the protein being analysed. In the variable selection method of Manavalan and Johnson [22], the selection of appropriate spectra can be made by the investigator. In any case, it is clear from Fig. 2 that the analysis of far-UV CD spectra can be undertaken more confidently when data are available in the range down to 180 nm or even beyond [23]. In practice, with many biological systems it can be difficult to acquire data of sufficiently high quality below 190 nm, or even 200 nm if other absorbing components are unavoidably present. (The intense radiation available from synchrotron sources offers one possible way of obtaining data at these low wavelengths [24], although such applications are still in their infancy). If data from wavelengths only above 200 nm are available, it is still possible to make fairly reliable estimates of the α-helical content of peptides and proteins from the ellipticities at 208 and 222 nm [25], the wavelengths corresponding to the characteristic double minima of the α-helix spectrum (Fig. 2). It should be pointed out, however, that the value of the mean residue ellipticity for an α-helix at 222 nm can be influenced to a significant degree by the length of the helix and the dynamical motion of the protein [20]. The analysis of far-UV CD spectra has been extensively reviewed [23,26–28].

2.1.2. The near-UV CD of proteins (aromatic amino acid dichroism)

The near-UV CD of proteins arises from the environments of each aromatic amino acid side chain as well as possible contributions from disulfide bonds, or non-protein cofactors which might absorb in this spectral region. Small model compounds of the aromatic amino acids exhibit CD spectra because the chromophore is linked to the nearby chiral α-carbon atom. In the case of proteins in their native states, the side chains of these amino acids will be placed in a variety of asymmetric environments characteristic of the tertiary structure of the folded protein. The analysis of near-UV CD spectra has been discussed in detail [1,29–31]. Each of the aromatic amino acids tends to have a characteristic wavelength profile: tryptophan, a peak close to 290 nm with fine structure between 290 and 305 nm; tyrosine, a peak between 275 and 282 nm (the fine structure at longer wavelengths may be obscured by that from tryptophan); phenylalanine, sharp fine structure between 255 and 270 nm [29]. Three factors tend to influence the intensities of aromatic cd bands: (i) the rigidity of the protein, with the more highly mobile side chains having lower intensities; (ii) interactions between aromatic amino acids, which are especially significant if the distance between them is less than 1 nm; (iii) the number of aromatic amino acids. Proteins with large numbers of such amino acids can have smaller than expected CD bands because of cancelling effects of positive and negative contributions, as in the case of aspartate carbamoyltransferase [32].

The contributions of individual aromatic amino acids to the near-UV CD spectrum of a protein can be assessed by examining the CD spectra of suitable mutant proteins. This approach was adopted by Craig et al. [33] to show that of the four tyrosines and one tryptophan in interleukin 1β, the major contributions were made by Trp-120 and Tyr-68. Freskgård et al. [19] used a similar approach to evaluate the contributions of each of the seven tryptophan residues in human carbonic anhydrase II, using a range of mutants in which each tryptophan had been replaced in turn. The fact that the mutant enzymes retained substantial activity (at least 34% of the wild type) indicated that the structures were likely to be very similar to wild-type enzyme. Each tryptophan was found to make distinctive contributions to the near-UV CD spectrum; when these were summed algebraically the resulting spectrum was qualitatively similar to that of the wild-type enzyme. In this case each of these side chains was found to make significant contributions not only to the near-UV but also to the far-UV CD (Fig. 3), varying between $+2500$ and -1500 deg. cm^2 $dmol^{-1}$ at different wavelengths. The contributions are particularly significant in the case of carbonic anhydrase where the ellipticity in the 220 to 230 nm region does not exceed -4000 deg. cm^2 $dmol^{-1}$. This example serves to emphasise the need to exercise care in the interpretation of far-UV CD spectra of proteins of low helical content, particularly when mutations involving aromatic amino acids are involved. In view of the various factors involved, it is hardly surprising that near-UV CD spectra of proteins

are not readily amenable to detailed interpretation in terms of tertiary structural features. Nevertheless, near-UV CD spectra can be very useful 'fingerprints' for comparisons of tertiary structures between related proteins, e.g., wild-type and mutant proteins (Section 2.3), and have been invaluable in studies of the 'molten globule' state of proteins (Section 3.5).

2.1.3. The near-UV / visible CD of proteins (cofactor / ligand dichroism)

CD in the near-UV, visible and near-IR can give a great deal of information on the environments of cofactors and other protein-bound ligands. In many cases, the free cofactor or ligand is optically inactive, but on binding to the protein becomes immobilised in an asymmetric environment, thus giving rise to CD signals in a characteristic absorption region. The CD spectra of complexes formed between carbonic anhydrases and a range of (intrinsically optically inactive) azosulfonamide inhibitors was used to explore detailed differences between the active sites of different isoenzymes [34]. Changes in the CD signal due to the bound cofactor or ligand can be used to report on the loss of integrity of the binding site during unfolding of the protein. For example, the GdnHCl-induced dissociation of the pyridoxal 5′-phosphate cofactor from phosphorylase *b* could be readily monitored by the loss of ellipticity at 335 nm [35]; this was complete at a concentration of denaturant (0.8 M) at which only very small changes in the secondary and tertiary structure of the enzyme had occurred as indicated by the far- and near-UV CD respectively. In a number of studies, the CD spectra of cytochrome *P*-450 enzymes over the range from 350 to 450 nm have been used to examine the hydrogen bonding and polarity characteristics of the active site haem group

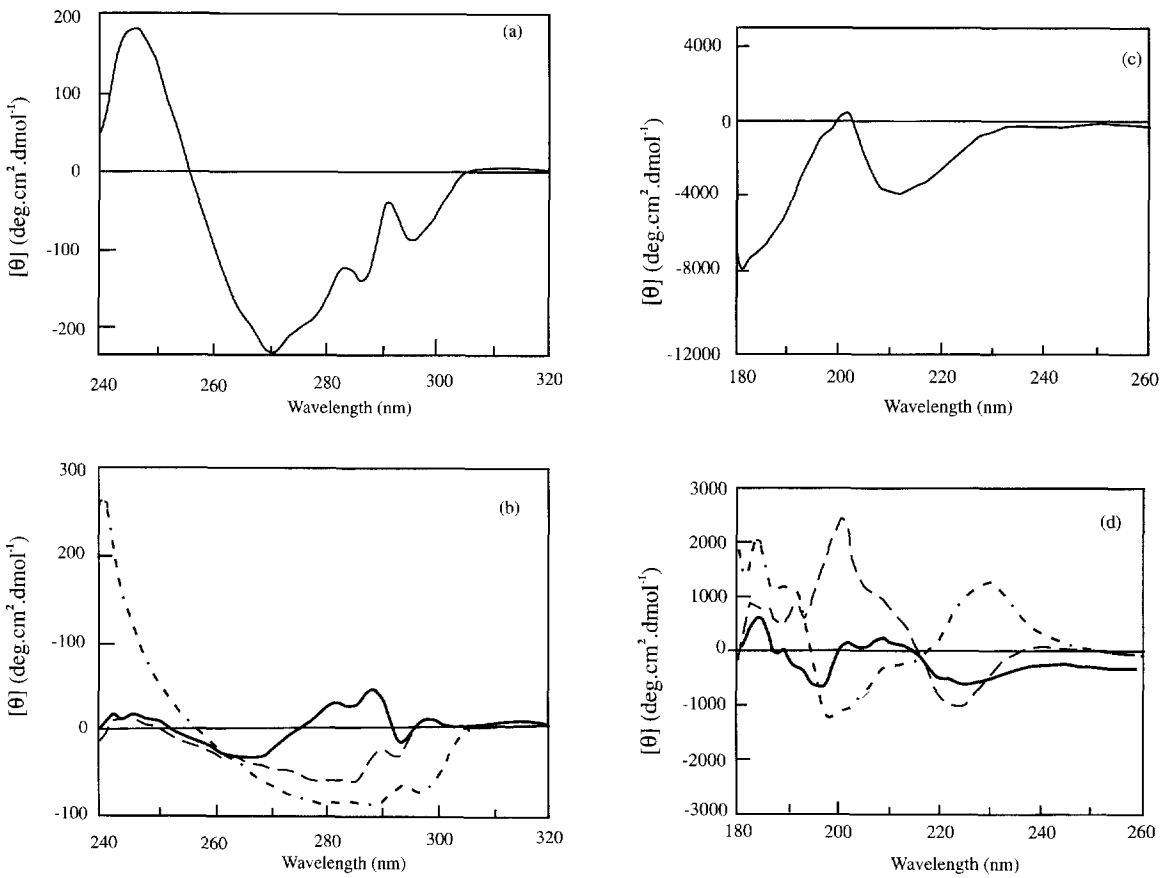

Fig. 3. CD spectra of carbonic anhydrase II. (a) and (c) Near-UV and far-UV CD spectra respectively of wild-type enzyme. (b) and (d) Contributions of Trp-5 (————), Trp-16 (- - -), and Trp-97 (– · –) to the near-UV and far-UV CD spectra respectively, as deduced from the spectra of appropriate mutants in which the tryptophan had been replaced by another amino acid [19].

[36] and the interactions between reductase and *P*-450 domains in the BM3 flavocytochrome system [37] (Fig. 4). The visible and near-UV CD spectra of solubilised light harvesting complexes from photosynthetic bacteria (such as *Rhodopseudomonas acidophila*) show a number of peaks due to the carotenoid and bacteriochlorphyll components) (Fig. 5). The spectrum in the 800 to 850 nm region provides a sensitive probe for interactions between neighbouring bacteriochlorophyll molecules in the complex [38]. Detailed analysis of the three dimensional structure of the light harvesting complex [39] shows that the carotenoid components are distorted by twisting from planarity, thus providing an explanation for the appearance of a CD signal (Fig. 5) from the otherwise optically inactive species (R.J. Cogdell, personal communication).

2.2. Experimental aspects of CD

In this section we shall give a brief outline of some of the important experimental aspects of obtaining CD data; this is important in appreciating the conditions under which reliable data can be obtained. Further details of these aspects can be found in a number of review articles [1,23,40,41].

The light source for most CD instruments is a xenon arc, which gives good output over the range of wavelengths (178 to 1000 nm) used for virtually all studies on proteins. It is necessary to flush the instrument with N_2 gas in order to remove O_2 from the lamp housing and the sample compartment so as (a) to prevent ozone formation and to minimise damage to the optical system, and (b) to allow measurements to be made below 200 nm. In a N_2 atmosphere, the lower wavelength limit is close to 180 nm; in order to record data below this wavelength, vacuum UV techniques or the use of high intensity synchrotron radiation sources [24] are necessary.

To obtain reliable CD data, it is important to pay attention to the instrument and to the sample. As far as the instrument is concerned, regular maintenance and calibration with a suitable chiral standard such as 1*S*-(+)-10-camphorsulfonic acid is essential. In view of the relatively small size of the signal a number of experimental parameters are adjusted in order to improve the quality of the data. These include: the time constant, the scan rate, the number of scans and the bandwidth. In most cases, suitable 'steady state' CD spectra in the near and far-UV can be obtained using a time constant of 2 sec, a scan rate of 10 nm/min, accumulation of two to four scans and a bandwidth of 2 nm or less [40].

It is important that the sample of protein should be

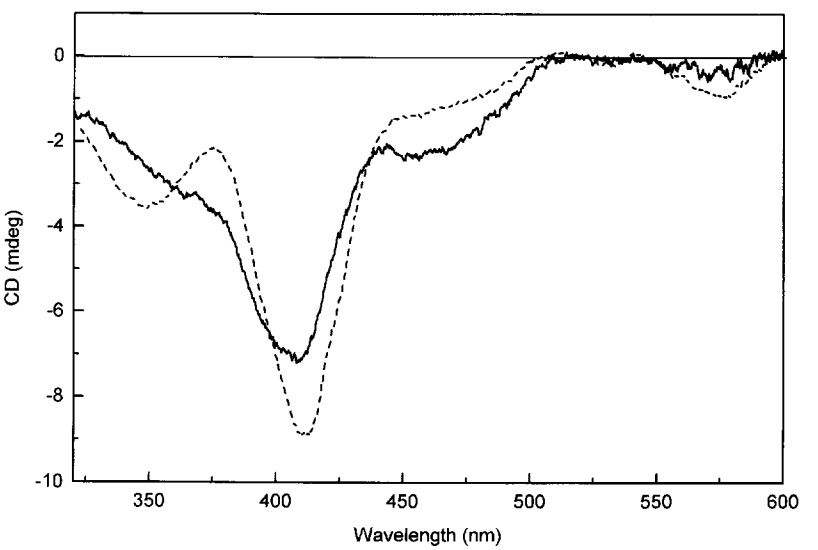

Fig. 4. The near-UV/visible CD spectrum of BM3 flavocytochrome. The solid line shows the CD spectrum of the intact flavocytochrome; the dashed line shows the CD spectrum of the stoichiometric mixture of the separately expressed *P*-450 and reductase domains of the enzyme [37].

homogeneous and should be freed of highly scattering particles by either centrifugation or passage through a suitable filter (e.g., 0.2 μm). The total absorbance of the sample should not exceed about one unit, otherwise the spectral noise will become excessive and, above a certain point, an automatic cutoff may operate leading to an apparent decline of the CD signal to zero. It is essential to minimise absorption due to other components (buffers, supporting electrolytes, solvents etc.) in the mixture. Most problems arise in the far-UV and the absorption characteristics of many commonly used buffers and other reagents have been extensively studied [40]. Phosphate, borate and low molarity (20 mM) Tris have low absorbances above 190 nm in cells of path length 0.1 cm or less; these buffers can between them give suitable coverage of pH values from 6 to about 9.5. Buffers which are appropriate for pH values from 4 to 6 usually contain carboxylate groups which have high absorbance below 200 nm. In such cases it is important to work with as dilute a buffer solution as is required to maintain the pH, and to run appropriate 'blank' CD spectra to ensure that the buffer components do not lead to excessive noise or other artifacts in the spectra. One point often overlooked is that Cl^- ions absorb strongly below 195 nm; thus NaCl is not recommended as a component of a buffer solution for CD work. Fluoride and sulfate salts are much more appropriate to maintain the ionic strength, if necessary.

In order to be able to estimate the secondary structure content of a protein from CD it is essential to know the protein concentration accurately (to within at least $\pm 5\%$, and preferably better). The

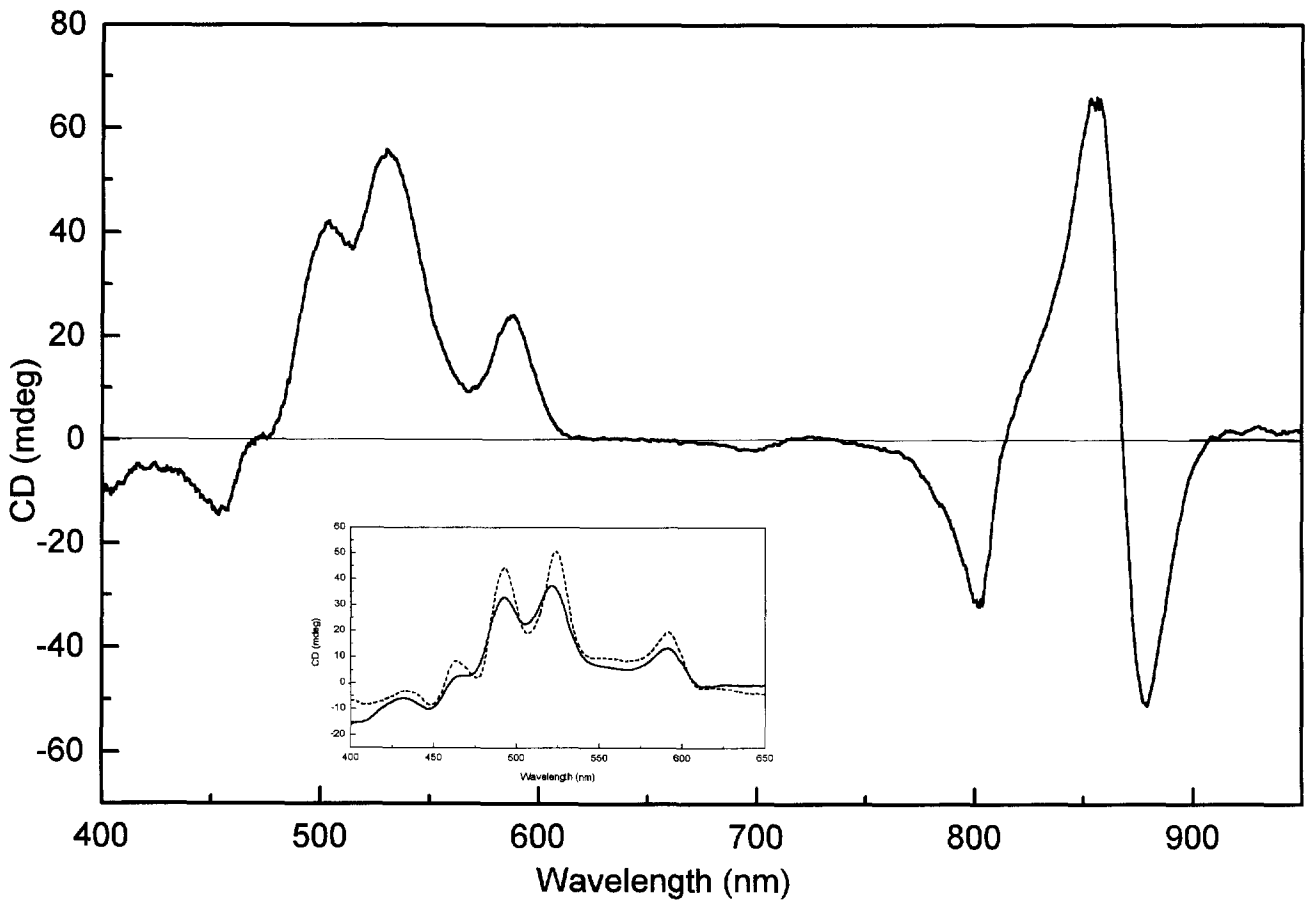

Fig. 5. The visible/near-IR CD spectrum of the light harvesting complex LH2 from *Rhodopseudomonas acidophila*. The complex was solubilised in buffer containing 0.1% LDAO. The inset shows spectra of a solubilised preparation to which 50% (v/v) glycerol had been added. The solid and dashed lines represent spectra recorded at 20°C and −150°C respectively.

most reliable methods are those which can be calibrated against absolute values such as A_{280} measurements calibrated with reference to amino acid analysis data [42]. For far-UV CD, where the dichroism of the peptide bond is being measured, it is the molar concentration of these bonds which is required; this can be obtained by dividing the concentration in mass terms (i.e., mg/ml) by the mean residue weight (strictly speaking, mass) or MRW. The MRW is obtained by dividing the molecular mass by $N-1$, where N is the number of amino acids in the polypeptide chain. For most proteins, the MRW is about 110, but can deviate significantly from this value if the amino acid composition is 'unusual' or if the protein undergoes extensive post-translational modifications.

When CD data are expressed in terms of absorbance, the units are those of the *difference in molar absorbance* $\Delta\epsilon = \epsilon_L - \epsilon_R$, i.e., $cm^{-1}\ M^{-1}$.

In terms of ellipticity, the *mean residue ellipticity* $[\theta]_{mrw,\lambda}$ at a wavelength λ is quoted in units of $deg.cm^2.dmol^{-1}$ and is given by:

$$[\theta]_{mrw,\lambda} = MRW\theta/10dc$$

where θ_λ is the observed ellipticity (degrees), d is the path length (cm) and c is the concentration (in units of g/ml).

Reservations have been expressed as to whether mean residue ellipticity is an appropriate unit for expressing the amplitude of the near-UV and visible CD signals of a protein, since only a small number of aromatic amino acids or prosthetic groups contribute to the CD signals in these regions. In such cases, it may be appropriate to express the data in terms of the molar concentration of intact protein, rather than of the repeating unit. At any wavelength, the numerical relationship between values in the two sets of units is:

$$[\theta]_{mrw} = 3298\,\Delta\epsilon$$

The amount of protein required for CD measurements can be gauged from the need to keep the absorbance less than about one unit. Typical cell path lengths for far-UV CD work are in the range 0.01 to 0.05 cm and protein concentrations are in the range 0.2 to 1 mg/ml. Depending on the design of the cell being used, the volume of sample required can range from about 1 ml to as little as 50 μl. Thus the amount of protein required to obtain an acceptable far-UV CD spectrum can be as little as 10 μg, but generally 100 to 500 μg sample is required in order to explore the optimal conditions for recording spectra. The CD signals in the near-UV and visible regions of the spectrum are much weaker than those in the far-UV, reflecting the much lower molar concentrations of the relevant chromophores compared with that of the peptide bonds. For measurements in these spectral regions it would be fairly typical to use a protein concentration of 0.5 to 2 mg/ml and a path length of 0.5 to 2 cm. The amounts of protein required for such experiments are thus of the order of several mg.

It should be noted that very useful information can be obtained by recording CD spectra at low temperatures (typically down to that of liquid N_2, i.e., 77 K) by use of mixed solvent systems or water/glycerol 'glasses'. The resolution can be considerably enhanced at low temperatures [1,29] (Fig. 5) and it may also be possible to examine the structures of individual 'frozen' conformational states which would rapidly interconvert at higher temperatures.

2.3. Strengths and weaknesses of CD as a structural technique

The purpose of this section is to evaluate CD as a structural technique, principally by comparison with X-ray crystallography and high-resolution NMR.

X-ray crystallography [43,44] requires crystals which diffract to sufficiently high resolution, and the availability of one or more isomorphous heavy atom derivatives; either of these requirements can take several weeks if not months. The X-ray structure of a protein, although providing detail at the atomic level, is essentially a static picture and can only afford limited insights into the dynamic aspects of protein structure, which may well be crucial to function.

The study of proteins by NMR requires high concentrations of sample (0.5–2 mM; i.e., 10–40 mg/ml for a 20 kDa protein) and a number of proteins are either not sufficiently soluble under these conditions (Section 3.5) or show a tendency to aggregate. For a complete determination of the structure, the individual resonances must be assigned and this requires multi-dimensional NMR techniques with (expensive!)

multi-isotope (^{1}H, ^{13}C, ^{15}N) labelling [45,46]. Using the highest magnetic field strengths currently available, full structural determinations are limited to proteins of molecular mass up to about 30 kDa. However, it should be emphasised that the NMR approach is capable of giving a wealth of information on both the structure and dynamics of proteins.

The principal advantages of CD derive from the speed and convenience of the technique. Compared with X-ray crystallography and NMR, CD measurements can be carried out rapidly; thus good quality spectra in the far and near-UV can each be obtained within 30 min. Far-UV CD studies, in particular, require only small amounts of material (Section 2.2) and since the technique is non-destructive, it is usually possible to recover most or all of the solution and hence conduct multiple experiments on the same sample. The ability to use cells of path lengths differing by several orders of magnitude means that it is relatively easy to study a very wide range of protein concentrations. For example, the changes in the far-UV CD spectrum of a 58-residue DNA-binding peptide derived from transcription factor GCN4 [47] over a 1000-fold range of concentrations could be interpreted in terms of a dimeric species with a high α-helical content (70–73%) dissociating into two unfolded polypeptide chains (Section 4.4). It was also possible to show that the model helix–turn–helix peptide (Section 4.4) did not undergo a concentration-dependent association over a 40-fold range of concentrations. In addition, CD studies can be performed over a wide range of experimental conditions, including pH, temperature and (with suitable attention to the requirement to avoid excess absorption or scattering of the incident radiation) in both solution and the solid state. Developments in instrumental design have now also made it possible to use stopped flow CD to study structural changes in proteins which occur on the timescale of a few milliseconds [48–50]. The low time constants required (typically of the order of 1 ms) lead to much greater levels of noise; however, the accumulation of data from a number of replicate experiments can produce results of good quality. The application of stopped flow CD measurements in both the far-UV and the near-UV has been of considerable value in helping to define structural changes occurring during some of the early steps in protein folding (see Sections 4.3 and 4.4).

The principal limitation of CD is that it only provides relatively low resolution structural information. Thus, although far-UV CD can give reasonably reliable estimates of the secondary structure content of a protein, in terms of the proportions of α-helix, β-sheet and possibly β-turns, it must be noted that these are overall figures and do not indicate which regions of the protein are of which structural type. (The value of the information from CD can be enhanced if used in conjunction with secondary structure prediction algorithms based on protein sequence [51]). At present, the development of CD theory is insufficient to allow detailed interpretation of the near-UV CD spectra in structural terms [1,29–31] and thus by itself CD is unable to give insights into the tertiary structure of a protein. However, even with these limitations, CD can still play a very useful role in structural work, especially when it is employed in a complementary fashion with other structural techniques. This is well illustrated by the studies of protein unfolding and folding (Sections 3 and 4), and is also evident in protein engineering studies. The secondary and tertiary structural characteristics of a range of mutant proteins can be assessed very rapidly by CD; this would allow the rapid selection of those mutants of altered structure which can then be characterised by the more detailed structural techniques. If the far-UV and near-UV CD spectra of the mutants are superimposable on those of the wild type protein, it can be concluded that the mutations have not resulted in any significant distortion of structure, as found for instance in inactive mutant forms of isocitrate lyase from *Escherichia coli* in which Cys-195 had been replaced by alanine or serine [52]. A slightly different case is illustrated in Fig. 6, which shows the CD spectra obtained for wild type and two mutant forms (R23A and R23K) of the type II dehydroquinase from *Streptomyces coelicolor*. Both mutants have less than 0.1% of the activity of wild type enzyme, consistent with the proposal that this arginine plays an essential role in catalysis [53]. The far-UV CD spectra (Fig. 6a) of the three proteins show that no changes in secondary structure have occurred in the mutants. In the near-UV CD (Fig. 6(b)), the spectrum in the 288–300 nm region is unaltered showing that the environment of the single tryptophan in each subunit of the enzyme has not been perturbed. However, there are significant

changes in the region of the spectrum (275–285 nm) characteristic of tyrosine residues, indicating that changes in the environment of one or more of the four tyrosines per subunit have occurred, consistent with the fact that a tyrosine which appears to be involved in catalysis (Tyr-28) is close to the essential arginine (Arg-23).

A further limitation of CD is that in general it gives no information on the quaternary structure of proteins, unless the association of subunits is accompanied by changes either in the secondary structure of the subunits (as in the case of the GCN4-derived peptide mentioned above) or in the environment of aromatic amino acid residues (this might be the case if these were at an interface, for example). Techniques such as gel permeation, light or X-ray scattering, or chemical cross-linking would be most commonly used in conjunction with CD measurements to explore the relationships between quaternary and secondary or tertiary structures [4].

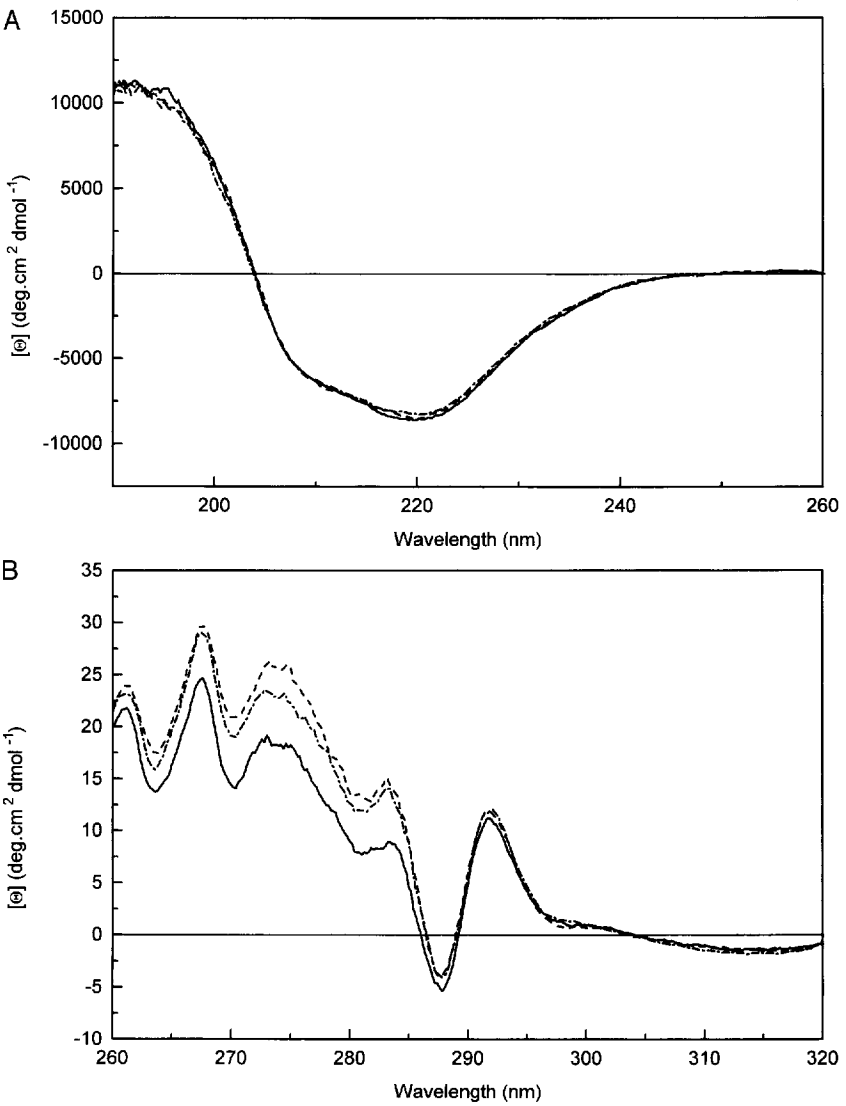

Fig. 6. CD spectra of type II dehydroquinase from *Streptomyces coelicolor*. (a) far-UV CD spectra, (b) near-UV CD spectra. (————), (- - -) and (– · –) represent the spectra of wild type, R23A and R23K mutants respectively [53].

3. Protein unfolding

3.1. Applications of cd in studies of protein unfolding

In this section the application of CD to study the overall characteristics of the unfolding of proteins, as well as to gain insight into the mechanisms of the processes will be outlined using systems of increasing complexity. Generally, the equilibrium aspects of the unfolding process have been studied, but kinetic data on unfolding can also be obtained.

3.2. Overall measures of protein stability

The unfolding of proteins is most commonly brought about either by heat or by the addition of denaturants such as urea or guanidinium chloride; in some special cases unfolding can be brought about simply by dilution, e.g., the GCN4-based peptide where a helical dimer dissociates into an unstructured monomer (Section 4.4). Changes in CD signals in either the far-UV or the near-UV can be used to monitor the loss of structure during unfolding; other spectroscopic parameters which can be measured are fluorescence and UV absorbance; alternatively thermal unfolding can be monitored by differential scanning calorimetry [54].

The information available from unfolding studies can be of a general type, such as the temperature or denaturant concentration at which 50% unfolding has occurred. These parameters can be used to provide comparative measures of stability either between different proteins (or mutants of the same protein) or in the case of a single given protein, the effect of modification or changes in environment. An example of the latter case is provided by the formation of a reduced imine complex at a single lysine side chain in each subunit of the dimeric enzyme type I dehydroquinase from *Salmonella typhi*. Formation of this derivative led to a remarkable increase in stability towards unfolding by GdnHCl, with the mid point being raised from 1.6 M to 4.5 M as monitored by the loss of the CD signal at 225 nm [55]. This increase in stability was also reflected in the T_m measured by differential scanning calorimetry from 56°C to 98°C. CD studies in the far and near-UV showed that the increase in stability does not involve any substantial changes in the overall secondary and tertiary structures of the protein. Instead it seems to be associated with a general restriction in the flexibility of the enzyme as indicated by a remarkable decrease in the accessibility of the single tryptophan in each subunit of the enzyme to the quenching agent succinimide [55].

3.3. Quantitative estimates of the stability of the folded state

In favourable cases it is possible to analyse the unfolding curve mathematically in order to derive estimates of the stability of the folded state relative to the unfolded state of the protein [14,56]. It is important to establish the characteristics of the unfolding curve, in order to see whether a simple two state model applies (i.e., N (native) $\rightleftharpoons$ U (unfolded)). Thus for instance the unfolding of barnase and RNase T1 by heat or denaturants obey the simple model as shown by (1) the shape of the unfolding curve, (2) the coincidence of the transition curves monitored by a number of different parameters, and (3) the equivalence of ΔH_{cal} and $\Delta H_{van't\ Hoff}$. In other proteins unfolding is more complex; thus in the case of the unfolding of bovine carbonic anhydrase B, the changes in the CD in the aromatic region run ahead of the changes in absorption spectroscopy or in optical rotatory dispersion (measuring secondary structure) [57]. It was suggested that unfolding of the protein occurred in a set of successive, reversible, unimolecular steps in which different types of structure were lost. A three state model was used to describe the urea-induced unfolding of the α-subunit of Trp synthase [58] monitored by changes in CD at 222 nm or absorbance at 286 nm. More recent studies [59] on this system have indicated that the properties of the equilibrium unfolding intermediate does not retain an intact N-terminal domain, but rather has the properties of a 'molten globule' (Section 3.5).

If a two state model is shown to apply, the free energy change (ΔG) (equal to $-RT\ln([U]/[N])$) is evaluated at each [denaturant]. It is of course difficult to gather accurate data outside the range from 10% to 90% unfolded protein (i.e., a range of ± 5.4 kJ/mol around the mid point value). In order to evaluate the stability of the folded form (i.e., $\Delta G°(H_2O)$), it is necessary to extrapolate the data from the transition range to a denaturant concentration of zero. The

methods of performing this (often quite lengthy extrapolation) have been discussed in detail [56]. A linear extrapolation model gives values for $\Delta G°(H_2O)$ which are in satisfactory agreement with those derived from calorimetric measurements for proteins such as cytochrome *c* and T4 lysozyme. However, for other proteins such as α-chymotrypsin more complex non-linear extrapolations based on alternative models [60] are required. Studies on a number of globular proteins give values of ΔG which are surprisingly small (in the range +20 to +50 kJ/mol) corresponding to the energy of only a few H-bonds; this is taken to reflect the balance between the large enthalpic and entropic terms in the folding process.

An analysis of the unfolding of barnase by urea [61] has explored the determination of $\Delta G°(H_2O)$ in detail. In the case of this protein there was a significant discrepancy (6.7 kJ/mol) between the $\Delta G°(H_2O)$ value obtained by linear extrapolation of the urea unfolding data (36.8 kJ/mol) and that obtained calorimetrically (43.5 kJ/mol). At 25°C, urea-induced unfolding of the protein has a mid point [urea] of 4.58 M. In order to obtain ΔG data over a range of lower [urea], thermal unfolding of the protein was studied in the presence of concentrations of urea up to 4.5 M. The thermal unfolding was monitored either by changes in ellipticity at 230 nm or by differential scanning calorimetry. Under all conditions tested the unfolding was found to be a fully reversible process. The results showed that significant deviation from linearity of the plot of ΔG vs. [urea] occurred at low [urea]; the experimental data could be fitted by including terms in [urea] and [urea]^2 in the equation for ΔG. When this improved equation was used the $\Delta G°(H_2O)$ value obtained by urea denaturation was in good agreement with that obtained calorimetrically.

3.4. The unfolding of helices in model peptides

As well as obtaining data on the overall stability of the folded structure of a protein, CD measurements can provide important information relating to the mechanism of the unfolding process (and hence on the mechanism of the reverse process of protein folding). Detailed analyses have been carried out of the unfolding of peptide helices [62] by urea in order to assess the contribution of this process to the overall energetics of protein unfolding. The peptides studied were based on repeating units of the sequence Ala-Glu-Ala-Ala-Lys-Ala and ranged in length from 14 to 50 residues. Such peptides are known to form monomeric helices in solution and the lack of hydrophobic cores and long-range tertiary interactions allow attention to be focused exclusively on helix unfolding. Unfolding was monitored by the change in ellipticity at 222 nm, which could be related to the helical content taking into account the chain length dependence of the CD signal for a helix. Over a wide range of urea concentrations there was found to be a very nearly linear dependence of ln *s* (where *s* is the equilibrium constant for helix formation) on [urea]. When the slope (*m*) of this plot (which is a measure of the dependence of the unfolding of helices on [urea]) can be multiplied by the number of helical residues in a variety of proteins, the values obtained agree within a factor of 2 with the actual *m* values obtained in the linear extrapolation procedures; not surprisingly the agreement is better for those proteins such as myoglobin and *trp* aporepressor with high helical contents. The results indicates that the interaction between urea and peptide groups accounts for a major part of the denaturing action of urea on proteins.

3.5. The 'molten globule' state of proteins

CD studies provided some of the first convincing demonstrations of a structural state of proteins intermediate between the native and the unfolded form. When incubated at pH 2, α-lactalbumin retains most of its native secondary structure as shown by the far-UV CD. However, under these conditions the near-UV CD is very much reduced (Fig. 7) indicating that native tertiary interactions are absent [63,64]. Studies of the Stokes radius (gel permeation) show that at pH 2 the molecule is compact, leading to the term 'molten globule' or 'compact intermediate'. In addition to the CD properties, a number of other characteristics of this type of structure have been described, including the lack of chemical shift dispersion in NMR and enhanced binding of the fluorescent dye ANS [65]. Molten globule states may represent important intermediates in protein folding pathways [66–69] (Section 4.4), and it has been suggested that they can be recognised by certain types of molecular

chaperones [8,70]. For these reasons, the molten globule state of a number of model proteins has been widely studied. Two different approaches have been used to show that the two domains of α-lactalbumin can function independently. By using variants of the protein in which selected cysteine residues were replaced by alanine and in which the residues corresponding to the β-sheet domain were replaced by linker glycines, it was shown that the α-helical domain forms a helical structure with a native-like tertiary fold. The corresponding variant containing the β-sheet domain was largely unstructured [71]; however, binding of Ca^{2+} to this domain led to the formation of native-like structure, a process which may well be the crucial event in the Ca^{2+}-induced folding of the molten globule state of the intact protein [72]. In a derivative of α-lactalbumin in which two of the four disulfide bonds are retained, the native structure and Ca^{2+}-binding ability of the β-sheet domain persist, whereas the α-helical domain is unfolded [73].

CD in conjunction with NMR and binding assays was used to provide evidence that the 14 kDa DNA recognition subunit of DNA methyltransferase from *Agmanellum quadruplicatum* was predominantly a molten globule in the absence of DNA [74]. This observation might provide an explanation of how such proteins could achieve specificity of DNA binding without requiring undesirably high affinity [64,75]. A number of advantages of CD as a structural technique are illustrated by this work, namely the ability to compare the structures of the protein in the absence and presence of urea (the protein was insufficiently soluble at the concentrations required for NMR in the absence of urea), the demonstration of secondary structure (far-UV CD) but not stable tertiary structure (near-UV CD), and finally the ability to focus on the structural changes in the DNA on binding since, at the wavelengths concerned (around 280 nm), the signal from the DNA in the complex is much greater than that from the protein.

3.6. The unfolding of a simple protein, apomyoglobin

CD was used in conjunction with amide exchange measurements to examine the unfolding pathway of apomyoglobin [76]. It had already been established that the unfolding of apoMb by acid did not follow a two-state model, but rather that there was an intermediate (termed I_1) which was maximally populated at pH 4.2. This intermediate had the characteristics of a molten globule, with about 70% of the helix content of native protein (as judged by the ellipticity at 222 nm) but lacking a fixed tertiary structure (as indicated by the low level of chemical shift dispersion in the ^{1}H-NMR spectrum). In the presence of relatively

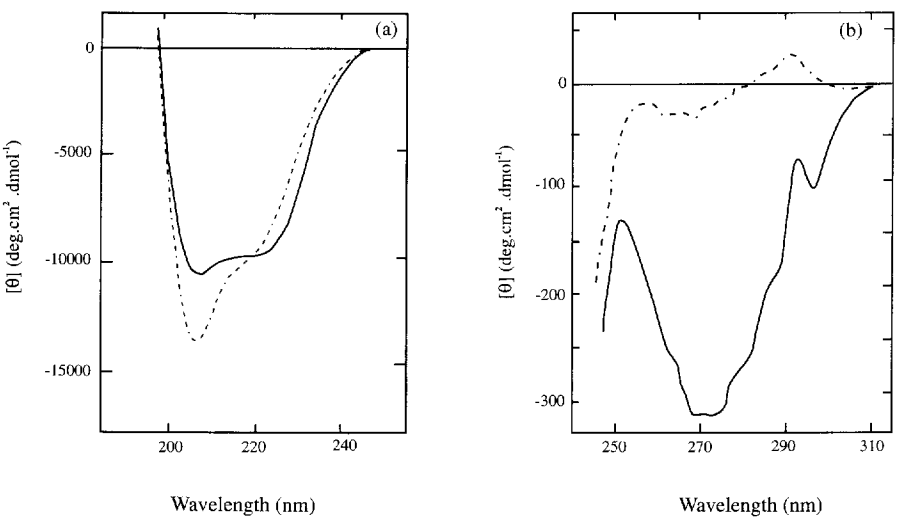

Fig. 7. CD spectra of α-lactalbumin. (a) and (b) represent far-UV and near-UV CD spectra respectively. The solid and dashed lines represent spectra obtained at pH 7 and pH 2 respectively [63,64].

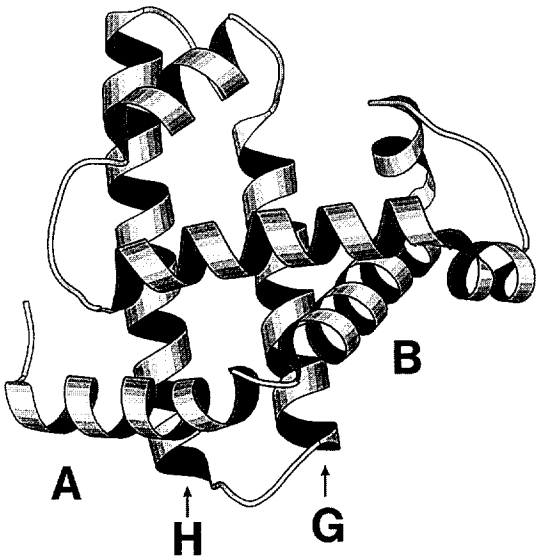

Fig. 8. The helical segments protected in the unfolding of apomyoglobin [76]. The diagram was produced from the coordinates for sperm whale myoglobin (5mbn.pdb) using the program MOLSCRIPT [118].

small concentrations (up to 5 mM) sodium trichloroacetate, the helicity (ellipticity at 222 nm) of the intermediate is increased (by about 10%) and a new intermediate state (termed I_2) is formed. (At pH 2, where apoMb is fully unfolded, addition of salts such as sodium chloride or sodium trichloroacetate leads to the formation of molten globule-type states, which can be shown by far-UV CD to differ in terms of their helical content [77]). CD was used to show that the I_1 and I_2 intermediates formed at pH 4.2 differed markedly in their stabilities towards denaturation by urea with mid point values being about 1.5 M and 3.9 M respectively. Indeed I_2 is only marginally less stable than native protein (the ΔG values are 24.7 kJ/mol and 27.6 kJ/mol respectively). From NMR amide protection experiments it is clear that the secondary structure in I_1 is associated with portions of the helices in the A-G-H subdomain of the protein (Fig. 8). In the I_2 intermediate it appears that these helices are further stabilised and, in addition, helix B has begun to form. In complementary work, NMR amide protection and stopped flow CD studies have been used to show that formation of these helices may well reflect early steps in the pathway of folding of the denatured protein [69].

3.7. The unfolding of more complex proteins

3.7.1. A two domain protein, phosphoglycerate kinase

Phosphoglycerate kinase has a single polypeptide chain of 415 amino acids, divided into two domains each of approx. 200 amino acids which mostly correspond to the *N*- and *C*-terminal halves of the sequence, although the *C*-terminal 15 amino acids (401–415) form an integral part of the *N*-domain. In order to assess the stabilities of the two domains of the enzyme from yeast, advantage has been taken of the fact that each can be produced in a folded form using recombinant DNA techniques [78]. Using the CD signal at 218 nm to monitor the loss of structure caused by GdnHCl, the unfolding of the intact enzyme and the isolated *N*- and *C*-domains was found in each case to be symmetric, monophasic and completely reversible. When the results were analysed according to the two-state model (Section 3.3), the stabilities ($\Delta G^{\circ}(H_2O)$ of the intact protein, the *N*-domain and the *C*-domain were found to be 32.6, 18.8 and 16.7 kJ/mol respectively. Thus in terms of thermodynamic stability, the intact protein behaves essentially as the sum of its component domains. However, there is evidence from the reactivity of the single cysteine (Cys-97) in the enzyme that the *N*-domain is somewhat more flexible when isolated compared with the intact protein, pointing to the presence of at least limited interactions between the two domains.

3.7.2. A two domain protein with multiple cofactors, flavocytochrome P-450

The flavocytochrome *P*-450 BM3 from *Bacillus megaterium* is a member of the cytochrome *P*-450 superfamily of monooxygenase enzymes. It is unusual in that it consists of two major domains each containing a redox centre (haem and flavin) linked covalently. The system offers a number of spectroscopic handles for CD. CD has been used to examine the unfolding of the protein and its domains by GdnHCl [79]. As measured by far-UV CD (225 nm) most of the loss of secondary structure occurs in the range 1 to 3 M GdnHCl. There is little difference between the unfolding profiles for the intact BM3 and its constituent domains, indicating that there is little mutual stabilisation of domains; a conclusion

consistent with the results of differential scanning calorimetry. CD in conjunction with absorption measurements in the near-UV and visible regions were used to show that loss of the redox cofactors from the isolated domains had begun to occur by 1 M GdnHCl and was complete by 2 M GdnHCl; at this concentration, loss of secondary structure had only occurred to the extent of about 60%. The overall activity of the enzyme or that of the isolated reductase domain is lost by about 0.4 M GdnHCl; which is correlated with the time-dependent removal of FMN from the reductase domain (FAD removal is complete by about 1.5 M GdnHCl).

3.7.3. A multisubunit protein, tyrosine hydroxylase

The study of the thermal unfolding of a more complex multisubunit enzyme is illustrated by the case of rat tyrosine hydroxylase [80], a tetramer of identical subunits each 56 kDa. Analysis of the far-UV CD spectrum of the enzyme at room temperature revealed helix and sheet contents of 55% and 22% respectively. As the temperature was raised, the CD signal at 220 nm was lost progressively preventing the measurement of a well defined T_m; since the unfolding was irreversible, it was not possible to obtain an estimate of the stability of the folded form as would be the case for less complex proteins (Section 3.3). Nevertheless, valuable information about the effects of a variety of factors (such as the presence of polyanions or phosphorylation by protein kinase A) on the stability of the enzyme could be obtained from CD measurements at a single elevated temperature.

4. Protein folding

4.1. Protein folding: general considerations

As discussed in Section 1.2, understanding the mechanism of protein folding is one of the major challenges of molecular biology. For nearly 30 years, since Levinthal pointed out that if the folding of a protein involved a sampling of all possible conformations available to it, the process would take an impossibly long time [81], it has been clear that protein folding must involve specific pathways. Amongst the stages which must occur during the folding of proteins are the formation of elements of recognisable secondary structure (helices, sheets, turns, etc.), the collapse of the extended polypeptide chain to form a more compact state in which non-polar side chains are largely buried away from the solvent, the formation of the distinct long-range interactions which characterise the native tertiary structure and are a prerequisite for the formation of specific binding and catalytic sites, and the association between subunits in oligomeric proteins [4,5,82,83]. Whether a single all-embracing model for protein folding exists is still a matter of debate (Section 4.5). The task for the experimental scientist is to employ a variety of techniques to gather structural and kinetic data which allow models for folding to be proposed and tested.

4.2. Experimental approaches to protein folding

By far the most detailed insights into the kinetics and mechanism of protein folding have come from studies of refolding, in which a mature protein is unfolded (usually by high concentrations of urea or GdnHCl; Section 3.2), and then refolding is initiated by rapid dilution of the denaturing agent. Thus a typical experiment might involve incubating a protein in 6 M GdnHCl, and performing a 60-fold dilution into buffer to initiate refolding. It is important to show that (a) the protein is completely unfolded prior to dilution and (b) that the residual concentration of denaturant after dilution (0.1 M in this example) has a negligible effect on the protein. For these purposes, CD in the near and far-UV together with studies of biological activity provide useful data; these techniques can also be used to measure the extent of the refolding of the protein.

In order to establish the mechanism of refolding it is necessary to study the time scale of the regain of structure from the denatured state, and to correlate these data with the rate of regain of biological activity. The regain of secondary structure is often monitored by far-UV CD and protection against amide hydrogen exchange (using NMR); that of tertiary structure by near-UV CD and fluorescence. The regain of quaternary structure can be monitored by rapid cross-linking with glutaraldehyde or by subunit hybridisation [4,83]; light scattering which would be useful in principle can be subject to artifacts caused by formation of aggregates. In view of the rates of

the various processes, it is usually necessary to use rapid kinetic methods (stopped or quenched flow) as well as manual mixing methods (dead time typically 10–15 sec) in order to probe as much of the refolding process as possible. Stopped flow methods typically have dead times in the range 2 to 10 msec, and involve dilutions of 11-fold (e.g., 6 M to 0.55 M GdnHCl, achieved by mixing of 1 volume of denatured protein with 10 volumes of buffer). Larger degrees of dilution using stopped flow techniques have been used but require considerable attention to factors such as the need to ensure rapid complete mixing; there can also be mechanical difficulties associated with driving syringes of very different sizes. The application of CD in conjunction with a number of other techniques has given some very useful insights into protein folding and allowed the formulation and testing of models, as outlined below.

4.3. The nature of early folding intermediates

Stopped-flow CD has been extensively used to examine the properties of the early intermediates in protein folding. In the first studies of this type [84], it was shown that during the refolding of denatured ferricytochrome *c* and β-lactoglobulin the native ellipticity in the far-UV was regained within the dead time of the instrument (estimated as 18 msec); by contrast the ellipticity in the visible region (ferricytochrome *c*) or the near-UV (β-lactoglobulin) was only regained over a period of minutes. These results were taken to indicate that within 20 msec an intermediate had formed in which a substantial proportion of native secondary structure was present; however, this intermediate had little stable tertiary structure. Similar results have been obtained for a number of other proteins including α-lactalbumin [85] and apomyoglobin [69]; these indicate that a high proportion of native secondary structure formation occurs within the dead time of the stopped flow CD experiment (which has been reduced in later instrumentation to about 5 msec). In a study of the refolding of native and mutant forms of staphylococcal nuclease, it was found that a transient kinetic intermediate formed within 10 msec [86]. This intermediate possessed only about 30% of the native ellipticity at 225 nm; in conjunction with hydrogen exchange experiments it was shown that this relatively low amplitude reflected formation of the 5-stranded β-sheet structural core of the N-terminal domain of the enzyme, with the helices (which contribute the bulk of the CD signal at 225 nm) formed at a later stage [87]. An extremely useful review of the use of stopped flow CD has been given by Kuwajima [50].

Further characterisation of the properties of early folding intermediates has involved the use of the fluorescent probe, ANS. For example studies of the β_2-subunit of tryptophan synthase [88] and the highly helical *trp* aporepressor protein [89] indicated that for both proteins a state had been formed within the dead times (13 and 5 msec, respectively) in which a high proportion of the native secondary structure had been formed (as indicated by the far-UV CD) and in which a significant proportion of the non-polar surface had been buried, as indicated by the large increase in ANS fluorescence compared with free ANS.

In the case of the β_2-subunit of tryptophan synthase, the ANS fluorescence passed through a maximum after about 1 sec and then slowly declined; this behaviour was interpreted in terms of the formation and decay of a 'molten globule' intermediate which bound ANS to a greater extent than either the unfolded or the native protein (Section 3.5). In the case of the *trp* aporepressor further increases were observed in the far-UV CD signal over a 5 sec period; these were accompanied by a further increase in ANS fluorescence. This increase was shown to be most probably due to the binding of the probe to the tryptophan binding site in the protein, since in the presence of the tryptophan analogue indole acrylic acid (which binds 100 fold more tightly than ANS), there is a marked decrease in ANS fluorescence after the initial dead time rise. Taken together with data on the changes in tryptophan fluorescence during refolding (where no burst phase is observed), a scheme for the refolding of the *trp* aporepressor was proposed involving later folding and association reactions of the intermediates to form the remaining secondary, tertiary and quaternary structure.

On a cautionary note it should be mentioned that Engelhard and Evans [90] have shown from studies on carbonic anhydrase and α-lactalbumin that experiments involving the use of ANS fluorescence to monitor the refolding of proteins have to be interpreted with caution. Markedly different kinetic patterns were observed in the cases when ANS was

present from the start of refolding from those when, using a pulse method, the probe was only introduced at chosen time intervals after refolding had been initiated. This indicated that interactions between the probe and the protein could interfere with the folding process itself. The conclusion reached by Engelhard and Evans was that states capable of binding ANS (probably resembling molten globules) were populated during refolding; the conventional experiments with ANS present throughout the folding did not allow accurate measurements of the kinetics of formation and decay of such states [90].

Some of the most detailed studies on protein folding to date have been those on the refolding of hen egg white lysozyme [17,91,92]; it should be noted that in these studies the four disulfide bonds of the enzyme are kept intact. In addition to stopped flow CD and fluorescence, pulsed amide hydrogen exchange labelling has been used to define those regions of the protein where secondary structure and tertiary interactions had formed at various stages of the refolding process. Taking advantage of the extensive NMR spectral assignment of lysozyme, it is possible to use the hydrogen exchange technique to monitor the fate of more than 50 amides (nearly half the total in the protein) during refolding. From these experiments it is clear that there are two classes of amides with distinct protection kinetics; about 50% of the amides are protected within 200 msec, the remainder are only protected after 1 sec. The finding that most of the rapidly protected amides are within the α-domain of lysozyme (containing the 4 α-helices), and the more slowly protected amides predominantly in the β-domain (containing the β-sheet) might be construed as indicating that the folding of these two domains is a sequential process. However, the fact that biphasic kinetics of protection are observed for both domains indicates that the folding process involves multiple pathways of folding with a variety of structured intermediates. Additional insights into the folding of lysozyme have been obtained from stopped flow CD measurements in the far- and near-UV (Fig. 9). The near-UV results (289 nm) show that the native ellipticity at this wavelength develops in a single kinetic process, with a rate very similar to the protection of amides in the β-sheet domain. Since the majority of the tryptophan side chains are in the α-domain, where the amides are

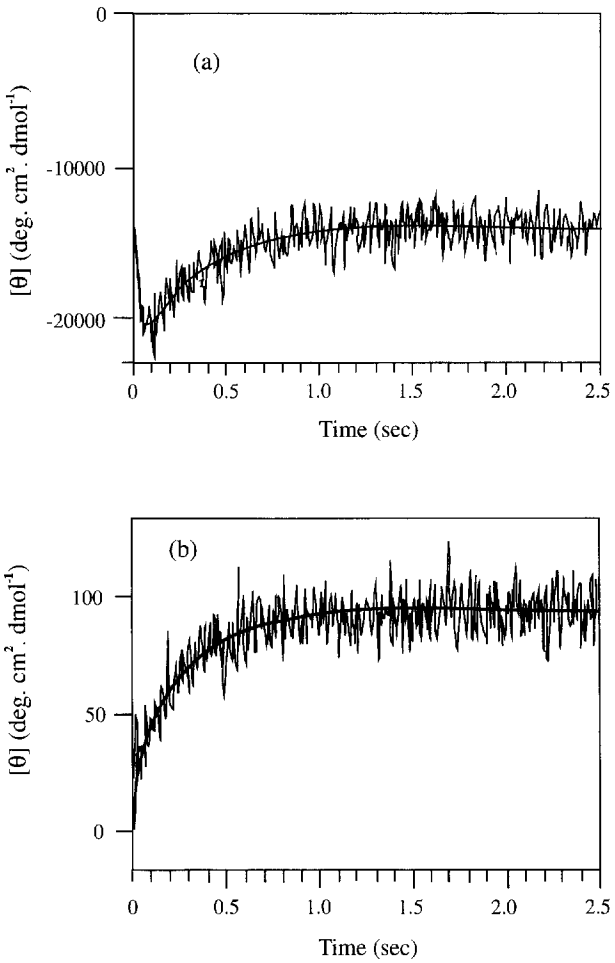

Fig. 9. The refolding of lysozyme after denaturation in GdnHCl [91,92]. (a) Refolding monitored by changes in the ellipticity at 225 nm. The mean residue ellipticities of denatured and folded protein samples are -1500 and -14000 deg.cm^2.dmol^{-1}, respectively. (b) Refolding monitored by changes in the ellipticity at 289 nm. The mean residue ellipticities of denatured and folded protein are 0 and 100 deg.cm^2.dmol^{-1} respectively.

protected faster, this indicates that the partially folded intermediate form of this domain has the characteristics of a molten globule, with the tryptophan side chains not yet having adopted their native orientations. The kinetics of the changes in CD signal at 289 nm are also very similar to those of the development of a binding site for a fluorescent derivative of chitobiose which is a competitive inhibitor of the enzyme.

By contrast the results from stopped-flow CD in the far-UV reveal a complex picture (Fig. 9). Within the dead time of the instrument (about 5 msec) about

80% of the native ellipticity at 225 nm has been regained. In a second phase, essentially complete after 80 msec, the signal at 225 nm increases to about 150% of the native value. The CD signal returns to the native value in a third phase (of half-time 300 msec). From these results it is clear that a large proportion of secondary structure is formed very quickly (within 2 msec) by which time there is only a very small degree of amide protection in the labelling exchange experiments. This suggests that such protection arises from the stabilisation of already formed helices (see below). The apparent 'overshoot' of the far-UV CD in the second phase of the process has been examined in detail [17]. The most likely explanation for the overshoot is contributions from side chains, of which the largest component comes from disulfide bonds, with only a small contribution from aromatic side chains. The sign of the CD signal of disulfide bonds in the far-UV is critically dependent on the –C–S–S–C– dihedral angle; thus the large negative ellipticity at 225 nm observed during the early stages of folding of lysozyme could arise from four disulfide bonds with dihedral angles close to 90° in the P (right-handed helix) chiral conformation [93] and has been observed in a number of disulfide cyclic peptides of various lengths containing a proline residue [94].

Guijarro et al. [95] have examined the relationship between the hydrogen exchange method and far-UV CD in assessing the degree of secondary structure formation in early folding intermediates. In addition to the work mentioned above on lysozyme, studies on the refolding of cytochrome *c* and interleukin 1β [96,97] had both indicated that far-UV CD revealed that a much greater proportion of secondary structure had been formed within 4 msec than that estimated by hydrogen exchange. The system examined by Guijarro et al. [95] was the 101 amino acid C-terminal proteolytic domain (F2) of *Escherichia coli* tryptophan synthase β chain which on the basis of its rate of folding to generate a native like far-UV CD signal and its ANS binding properties had been previously proposed to resemble a typical early intermediate in protein folding [98]. CD in conjunction with FTIR showed that 45% of the residues of F2 were involved in secondary structural elements. However, the NMR spectrum of F2, particularly in terms of line broadening, absence of structure induced chemical shifts and absence of NOEs in the amide region, indicated that the isolated F2 adopts a number of structures which are in rapid equilibrium with each other. The hydrogen exchange protection factors, estimated from one-dimensional NMR are less than 60 for the majority of protons and probably closer to 10. Low protection factors of these magnitudes would account for the lack of protection seen at the very early time points in protein folding.

The above studies have shown that a compact molten globule-like intermediate possessing the majority of the native secondary structure is formed on the millisecond time scale, i.e., within the dead time of stopped flow instruments. The relationship between the rate of hydrophobic collapse and the rate of secondary structure formation has until recently remained unresolved. However, studies by Agashe et al. [99] on the refolding of barstar (an 89 amino acid protein which is a powerful inhibitor of barnase) after GdnHCl-induced unfolding have allowed the situation to be resolved at least for this protein. The compactness of the molecule was measured by the efficiency of energy transfer between the tryptophan side chains (there are three distributed at both ends of its bundle-like structure) and dansyl groups attached to the cysteine side chains. Using this approach, it was possible to show that the collapse of the polypeptide chain to a compact form had occurred within 4 msec after the start of refolding of barstar. This burst intermediate was able to bind ANS, but showed essentially none of the secondary structure (far-UV CD) or tertiary structure (near-UV CD or intrinsic tryptophan fluorescence) of the native protein. Both secondary and tertiary structures were regained in slower biphasic processes. The generality of the conclusions reached for barstar remains to be established.

4.4. Structural elements in protein folding

Intermediates in protein folding are intrinsically difficult to study since they may be present in only small concentrations and for short periods of time. An alternative approach has been to study the behaviour of relatively small model peptides which have been designed to form isolated elements of secondary structure. It is then possible to study the structures and rates of formation of these peptide systems without the complications caused by subse-

quent processes. Far-UV CD has been probably the most widely used technique to study the structure of such model peptides, along with FTIR and NMR (for a review see [100]). One of the first studies dealt with a peptide consisting of the N-terminal 13 amino acids of bovine ribonuclease, which was shown by far-UV CD to be capable of forming an α-helix in aqueous solution. The extent of helix formation was greater at low temperature, at high ionic strength and in the presence of organic solvents such as trifluoroethanol [101]. The apparent discrepancies between the determinations of secondary structure by NMR and CD may in part derive from the fact that the CD signal is based on contributions from all chromophores throughout the peptide and thus provides a measure of the average helix character of the peptide. By contrast, NMR identifies secondary structure on a residue by residue basis in terms of strong NOEs. The strength of CD signals is also reduced in shorter helices and by 'fraying' of the ends of the helices [100].

The formation of β-sheet structures in short peptides is much less well documented, but an example is found in a peptide based on an 18-residue sequence of bovine S-antigen. On the basis of CD and FTIR the peptide at pH values between 4 and 9.5 adopts an intermolecular β-sheet structure stabilised by salt bridges between the peptide chains [102].

Further developments of this type of study have involved linking such elements of structure together. Oas and Kim [103] designed a model for a proposed intermediate in the refolding of bovine pancreatic trypsin inhibitor (BPTI) (Section 4.5) composed of two peptides (16 and 14 residues in length) linked by a disulfide bond. This new peptide was shown by CD to have a folded structure and detailed NMR measurements showed that it possessed the secondary and tertiary structural features found in the corresponding regions of the intact BPTI molecule.

Fezoui et al. [104] designed a 38-residue peptide in which secondary structural elements could be linked; this could represent a model for a helix–turn–helix folding intermediate. The peptide had the following sequence and is depicted in Fig. 10 in the helical wheel format:

succinyl-DWLKARVEQELQALEAR**GTDS**NAE-LRAMEAKLKAEIQK-NH_2

The peptide was designed to form two 17-residue helices, with a four-residue linker (turn) sequence, which is shown above in bold. The side chains of the hydrophobic residues in the interface between the helices (underlined) were arranged to interact in a manner similar to that observed in coiled-coil peptides. The turn residues (GTDS) were chosen as consensus residues from β-turn regions involving four consecutive amino acids in protein structures. G and S act as strong helix-breaking and good N-capping residues.

Gel permeation chromatography showed that the designed peptide was monomeric in solution. No differences in the ellipticity at 222 nm were observed over the concentration range from 5 μM to 200 μM, or in the chemical shifts in NMR over the range from 200 μM to 5 mM, indicating that over the entire range, the peptide remained in a monomeric form. The CD spectrum showed that the peptide had a high proportion of α-helix, estimated at around 60%. This value is somewhat lower than expected ((34/38) $\times$ 100%, i.e., 89%) which might reflect the end fraying effect referred to above. By monitoring the CD spectrum, the peptide was shown to undergo a gradual unfolding as the temperature was raised from 5°C to 90°C, but residual structure was still present at the high temperature. The process was fully reversed as the temperature was decreased. The lack of a sharp unfolding transition is characteristic of a number of coiled-coil states and the molten globule states of certain proteins. NMR studies showed that the helices were formed as expected and associated with each other in the expected anti-parallel fashion (Fig. 10). The residues GTDS were also shown to form a turn structure. Thus by all the criteria assessed the peptide had the expected structural properties and thus resembled a protein folding intermediate. Further studies of this model peptide system, including studies of the rate of formation of, and association between, the helices could give insights into the interplay between secondary structure formation and longer range interactions in protein folding.

Minor and Kim [105] have examined the extent to which the tendencies of individual amino acids or sequences of amino acids reflect local preferences or longer range factors. Although a number of studies have examined the local aspect, the longer range factors are less well understood. An 11-residue sequence, AWTVEKAFKTF, (termed the 'chameleon'

sequence) was designed which folds as an α-helix when in one position but as a β-sheet when in another position of the 57-residue peptide (the IgG-binding domain of protein G (GB1). The sequence was designed to replace α-helix residues 23–33 (AATAEKVFKQY) or β-sheet residues 42–52 (EWTYDDATKTF) of GB1, to yield chameleon-α or chameleon-β respectively. In each case, the hydrophobic nature of the residues which constitute the interface between each of these secondary structure elements and the core of the peptide was preserved. The NMR spectra (NOE patterns) of chameleon-α and chameleon-β were very similar to those of GB1, indicating that the structures of all three proteins were essentially identical, and in particular that the 11-residue had adopted the expected α-helical or β-sheet conformation in the two chameleon proteins. In addition the two chameleon proteins displayed the expected biological activity (i.e., binding to Fc) but with 10–30-fold reduced affinity compared with GB1, probably as a result of mutation of important interface side chains. Far-UV CD (ellipticity at 218 nm) was used to monitor the thermal unfolding of the chameleon proteins. Both showed reversible cooperative unfolding with mid points at 62°C and 40°C for chameleons-α and -β respectively, characteristic of compact single domain globular proteins. An 11-residue peptide consisting of the chameleon sequence (Ac-AWTVEKAFKTF-NH_2) was shown by CD and NMR to be unfolded in solution. This clearly indicates showing that the secondary structure formed by the sequence in the chameleon proteins must be specified by long-range, rather than local interactions and emphasises the importance of the tertiary structural context in protein folding. A similar type of conclusion could be drawn from work on three peptide fragments (16 to 18 amino acids in length) from the protein β-lactoglobulin [106]. In the structure of the (predominantly β-sheet) protein, two of the fragments (11–28 and 61–77) correspond to β-strands, whereas the third (127–142) forms an α-helix. In aqueous solution all three peptides showed little secondary structure, but the addition of high concentrations of trifluoroethanol gave rise in each case to marked α-helical structure as shown by CD and NMR measurements. Clearly the results obtained on small fragments of proteins in the absence or presence of structure-promoting solvents cannot always be extrapolated to the secondary structure of the native protein.

Peptides based on leucine zipper proteins represent simple model systems for studying the folding and association of proteins [107]. The synthetic 33-residue peptide GCN4-p1, derived from the yeast transcriptional activator GCN4 forms a stable bimolecular coiled-coil structure, with the ellipticity at 222 nm indicating that there is essentially 100% helix. The loss of the CD signal at 222 nm was used to monitor the GdnHCl-induced unfolding of this peptide at a number of different peptide concentrations from 6 to

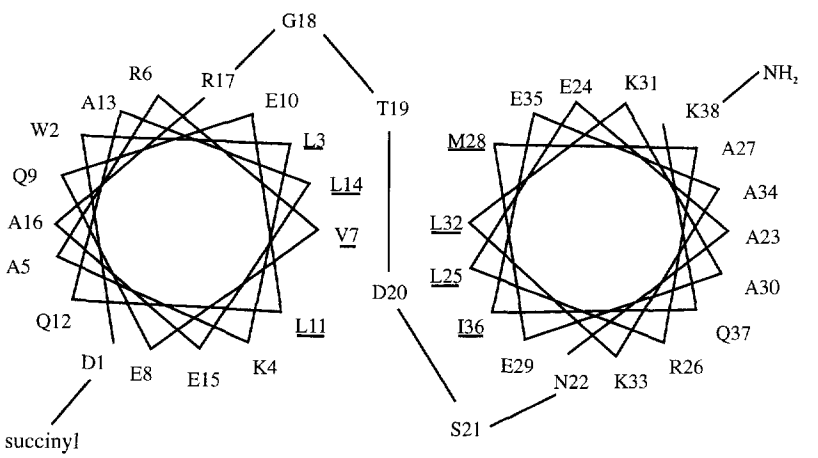

Fig. 10. Representation of the model helix–turn–helix 38-residue peptide [104] in the helical wheel format. The wheels run in opposite directions to reflect the fact that the helices are anti-parallel to each other. The helices have been depicted so that the hydrophobic side chains (underlined) form the interface between them.

37.5 μM (expressed as monomers). From the results it was clear that the midpoint of the transition increased with peptide concentration (from 1.89 M to 2.58 M GdnHCl over the range stated). These results were analysed successfully in terms of a model in which a folded dimer was in equilibrium with two unfolded monomers ($N_2 \rightleftharpoons 2U$); at 5°C and pH 7.0, the value of $\Delta G°(H_2O)$ is 43.8 kJ/mol dimer. The transition was found to be reversible on dilution of the GdnHCl. Using stopped-flow CD measurements at 222 nm (data collected over a period of 2 sec), the rate of refolding was found to be strongly dependent on the peptide concentration. Using a 10-fold dilution of GdnHCl (5 M to 0.5 M), the data over the stated range of peptide concentrations was found to fit the model $2U \rightarrow N_2$, with a second order rate constant of 2.2×10^5 M^{-1} s^{-1}. There was no significant regain of the CD signal within the dead time of the instrument (estimated as 5 msec). It is reasonable to suppose that some partially folded monomeric intermediate must be formed as a prerequisite for productive dimerisation. This might, for example, involve helix formation to generate hydrophobic interfaces required for association. However, since the data are very well fitted by the two state model, this indicates that partially folded intermediates are not sufficiently stable relative to the unfolded form to be significantly populated and hence detectable by CD during folding; this would be consistent with the rather low value of the second order rate constant which is some 3 orders of magnitude lower than would be expected for diffusion controlled association of peptides of this size.

4.5. General comments on protein folding models

From studies of the type described in Sections 4.3 and 4.4 it is clear that CD has an important role to play in providing structural and kinetic information about the pathway(s) involved in protein folding. Measures of overall secondary structure and the formation of native-like tertiary interactions from the far- and near-UV CD respectively have been complemented by a variety of other techniques which can examine the structural state of intermediates on a residue by residue basis (such as hydrogen-exchange NMR methods), or which can examine the formation of distinct local structural features such as substrate or inhibitor binding sites or epitopes. From all this work, various proposals concerning plausible models for protein folding can be made [1,5,108,109], though it is still unresolved whether any single model can apply universally.

It is now clear that some (generally small) proteins can fold from a denatured state extremely rapidly with no detectable kinetic intermediates. The cold shock protein CspB from *Bacillus subtilis* is a 67 residue protein whose tertiary structure consists of a 5-stranded β-barrel. Both the equilibrium unfolding and kinetic refolding behaviour (as monitored by changes in the fluorescence of the single tryptophan) conform perfectly to a two state model [110]. The small 4-helix bundle protein acyl Coenzyme A binding protein shows analogous folding behaviour [111].

However, in probably the majority of proteins studied to date, the process of refolding appears to involve the participation of one or more kinetic intermediates. In such cases, two main types of model have been proposed. One such model is a 'framework' one, in which acquisition of the final structure of a protein is seen in terms of a sequence of events beginning with very early formation of secondary structural elements which then coalesce into a compact molten globule-like form. The formation of native tertiary interactions (and with it the biologically active molecule) develops slowly from this compact molten globule. An alternative 'hydrophobic collapse' model places greater emphasis on an early collapse of the extended polypeptide chain to bury the non-polar side chains. This reduction in conformational space would be presumed to aid the subsequent formation of the native secondary and tertiary structures; however, in the case of cytochrome *c*, it has been shown that the 'collapse' process is likely to be the rate-limiting step in folding, and that once collapse has occurred folding occurs rapidly in an apparent two-step process [112]. In all models, formation of quaternary structure is seen as a very late event. The subunit recognition sites must be in essentially their native-like state before association occurs, since the latter process is known to be highly specific [4,5,83].

In those cases where observations show that compact states with substantial secondary structure are formed very early, usually within the dead time of a stopped-flow experiment, it is difficult to resolve the

question of whether a 'framework' or 'collapse' folding model applies. In the case of the refolding of barstar (Section 4.3), it is clear that hydrophobic collapse precedes formation of significant secondary structure [99], but the generality of this conclusion remains to be established. It is possible that the nature of the initial event(s) in the folding of these proteins may be explored by using conditions (such as higher residual denaturant concentrations) at which the relative stabilities and rates of formation of various intermediates in the pathway are altered.

It is often assumed that intermediates in a folding pathway occur in a linear sequence. While this may be true for many proteins e.g., barnase and chymotrypsin inhibitor 2 [113], the work on lysozyme and the *trp* aporepressor (Section 4.3) indicates that folding can proceed via a number of parallel pathways each with distinct intermediates. One possible source of heterogeneity in folding pathways is the slow rate of *cis* ⇌ *trans* isomerisation of Xaa-Pro peptide bonds, as in the examples of ribonucleases A and T1 [114]. In other cases, of which the classic example is the oxidative refolding of BPTI, some intermediates act as kinetic traps. Thus the intermediate in which two (30–51 and 14–38) of the three native disulfide bonds have been formed accumulates during refolding, but does not lead directly to the native protein. Instead slow intramolecular rearrangements occur to give the (30–51,5–55) intermediate which yields native protein by completion of the 14–38 disulfide bond [115]. Kinetic traps have also been observed in the folding of ribonuclease T1 and a lysozyme derivative with one of the four disulfide bonds broken [108].

The final paragraph of this review must address the question of the relevance of the studies in vitro on the refolding of (mainly) small proteins to the process of protein folding during biosynthesis. Clearly the latter process involves several new factors such as the possibility of co-translational folding and the effect of post-translational modifications such as disulfide bond formation, glycosylation and proteolytic processing, as well as translocation to a variety of intra- and extracellular destinations. It is not surprising that in the cell a range of additional factors such as chaperone proteins (e.g., those of the Hsp70 and chaperonin families) and enzymes such as protein disulfide isomerase [116] and peptidyl proline isomerase [114] are now known to be involved in ensuring that protein folding is both rapid and efficient [6–8]. Analysis of these more complex processes will build on the insights obtained from the study of the refolding of small model proteins, but will require the application of an array of sophisticated cell biological techniques. CD is probably of relatively limited use in the analysis of folding processes involving complex mixtures of proteins, unless it is possible to focus on distinctive signals from the protein of interest. While this is unlikely to be feasible in the far-UV, it might well be so in the near-UV if some of the other components lacked aromatic amino acids, for example. This could be the case for complexes involving the chaperonin GroEL, which lacks tryptophan residues [117].

Acknowledgements

We wish to thank the Biotechnology and Biological Sciences Research Council for financial support of the CD facility, Professor Richard Cogdell and Dr. Tino Krell for helpful discussions, and Dr. Tim Wess for help in producing Fig. 8.

References

[1] Adler, A.J., Greenfield, N.J. and Fasman, G.D. (1973) Methods Enzymol. 27, 675–735.
[2] Woody, R.W. (1995) Methods Enzymol. 246, 34–71.
[3] Fasman, G.D. (ed.) (1996) Circular Dichroism and the Conformational Analysis of Biomolecules 738 pp., Plenum Press, New York.
[4] Jaenicke, R. (1987) Prog. Biophys. Mol. Biol. 49, 117–237.
[5] Jaenicke, R. (1991) Biochemistry 30, 3147–3161.
[6] Gething, M.-J. and Sambrook, J. (1992) Nature 355, 33–44.
[7] Hlodan, R. and Hartl, F.U. (1994) in Mechanisms of Protein Folding (Pain, R.H., ed.), pp. 194–228, Oxford University Press, Oxford.
[8] Hartl, F.U. (1996) Nature 381, 571–580.
[9] Marston, F.A.O. (1986) Biochem. J. 240, 1–12.
[10] Thatcher, D.R. and Hitchcock, A. (1994) in Mechanisms of Protein Folding (Pain, R.H., ed.), pp. 229–261, Oxford University Press, Oxford.
[11] Ramage, R., Green, J., Muir, T.W., Ogunjobi, O.M., Love, S. and Shaw, K. (1994) Biochem. J. 299, 151–158.
[12] Thomas, P.J., Qu, B.-H. and Pedersen, P.L. (1995) Trends Biochem. Sci. 20, 456–459.
[13] Sifers, R.N. (1995) Struct. Biol. 2, 355–357.

[14] Pace, C.N. (1990) Trends Biotechnol. 8, 93–98.
[15] Kosen, P.A., Creighton, T.E. and Blout, E.R. (1981) Biochemistry 20, 5744–5760.
[16] Takagi, T. and Ito, N. (1972) Biochim. Biophys. Acta 257, 1–10.
[17] Chaffotte, A.F., Guillou, Y. and Goldberg, M.E. (1992) Biochemistry 31, 9694–9702.
[18] Woody, R.W. (1994) Eur. Biophys. J. 23, 253–262.
[19] Freskgård, P.-O., Mårtensson, L.-G., Jonasson, P., Jonsson, B.-H. and Carlsson, U. (1994) Biochemistry 33, 14281–14288.
[20] Hirst, J.D. and Brooks, C.L.,III (1994) J. Mol. Biol. 243, 173–178.
[21] Provencher, S.W. and Glöckner, J. (1981) Biochemistry 20, 33–37.
[22] Manavalan, P. and Johnson, W.C., Jr. (1987) Anal. Biochem. 167, 76–85.
[23] Johnson, W.C., Jr. (1990) Proteins Struct. Funct. Genet. 7, 205–214.
[24] Sutherland, J.C. (1996) in Circular Dichroism and the Conformational Analysis of Biomolecules (Fasman, G.D., ed.), pp. 599–633, Plenum Press, New York.
[25] Barrow, C.J., Yasuda, A., Kenny, P.T.M. and Zagorski, M.G. (1992) J. Mol. Biol. 225, 1075–1093.
[26] Yang, J.T., Wu, C.-S.C. and Martinez, H.M. (1986) Methods Enzymol. 130, 208–269.
[27] Greenfield, N.J. (1996) Anal. Biochem. 235, 1–10.
[28] Venyaminov, S.Yu. and Yang, J.T. (1996) in Circular Dichroism and the Conformational Analysis of Biomolecules (Fasman, G.D., ed.), pp. 69–107, Plenum Press, New York.
[29] Strickland, E.H. (1974) Crit. Rev. Biochem. 2, 113–175.
[30] Kahn, P.C. (1979) Methods Enzymol. 61, 339–378.
[31] Woody, R.W. and Dunker, A.K. (1996) in Circular Dichroism and the Conformational Analysis of Biomolecules (Fasman, G.D., ed.), pp. 109–157, Plenum Press, New York.
[32] Griffin, J.H., Rosenbusch, J.P., Weber, K.K. and Blout, E.R. (1972) J. Biol. Chem. 247, 6482–6490.
[33] Craig, S., Pain, R.H., Schmeissner, U., Virden, R. and Wingfield, T.P. (1989) Int. J. Pept. Protein Res. 33, 256–262.
[34] Coleman, J.E. (1968) J. Biol. Chem. 243, 4574–4587.
[35] Price, N.C. and Stevens, E. (1983) Biochem. J. 213, 595–602.
[36] Andersson, L.A. and Peterson, J.A. (1995) Biochem. Biophys. Res. Commun. 211, 389–395.
[37] Munro, A.W, Lindsay, J.G., Coggins, J.R., Kelly, S.M. and Price, N.C. (1994) FEBS Lett. 343, 70–74.
[38] Cogdell, R.J. and Scheer, H. (1985) Photochem. Photobiol. 42, 669–689.
[39] McDermott, G., Prince, S.M., Freer, A.A., Hawthornethwaite-Lawless, A.M., Papiz, M.Z., Cogdell, R.J. and Isaacs, N.W. (1995) Nature 374, 517–521.
[40] Martin, S.R. (1996) in Proteins Labfax (Price, N.C., ed.), pp. 195–204, Bios Scientific Publishers, Oxford.
[41] Johnson, W.C., Jr. (1996) in Circular Dichroism and the Conformational Analysis of Biomolecules (Fasman, G.D., ed.), pp. 635–652, Plenum Press, New York.
[42] Price, N.C. (1996) in Enzymology Labfax (Engel, P.C., ed.), pp. 34–41, Bios Scientific Publishers, Oxford.
[43] McRee, D.E. (1993) Practical Protein Crystallography, Academic Press, London.
[44] Drenth, J. (1994) Principles of Protein Crystallography, Springer, Berlin.
[45] James, T.L. and Oppenheimer, N.J. (eds.) (1994) Nuclear Magnetic Resonance Part C, Methods Enzymol. Vol 239, 813 pp., Academic Press, San Diego, CA.
[46] Whitehead, B. and Waltho, J.P. (1996) in Proteins Labfax (Price, N.C., ed.), pp. 205–216, Bios Scientific Publishers, Oxford.
[47] Weiss, M.A., Ellenberger, T., Wobbe, C.R., Lee, J.P., Harrsion, S.C. and Struhl, K. (1990) Nature 347, 575–578.
[48] Luchins, J. and Beychok, S. (1978) Science 199, 425–426.
[49] Pflumm, M., Luchins, J. and Beychok, S. (1986) Methods Enzymol. 130, 519–534.
[50] Kuwajima, K. (1996) in Circular Dichroism and the Conformational Analysis of Biomolecules (Fasman, G.D., ed.), pp. 159–182, Plenum Press, New York.
[51] Ribas de Pouplan, L, Atrian, S., Gonzalez-Duarte, R., Fothergill-Gilmore, L.A., Kelly, S.M. and Price, N.C. (1991) Biochem. J. 276, 433–438.
[52] Robertson, A.G.S. and Nimmo, H.G. (1995) Biochem. J. 305, 239–244.
[53] Krell, T., Horsburgh, M.J., Cooper, A., Kelly, S.M. and Coggins, J.R. (1996) J. Biol. Chem. 271, 24492–24497.
[54] Gilletto, A. and Pace, C.N. (1996) in Proteins Labfax (Price, N.C., ed.), pp. 233–239, Bios Scientific Publishers, Oxford.
[55] Moore, J.D., Hawkins, A.R., Charles, I.G., Deka, R., Coggins, J.R., Cooper, A., Kelly, S.M. and Price, N.C. (1993) Biochem. J. 295, 277–285.
[56] Pace, C.N. (1986) Methods Enzymol. 131, 266–280.
[57] Wong, K.-P. and Tanford, C. (1973) J. Biol. Chem. 248, 8518–8523.
[58] Matthews, C.R. and Crisanti, M.M. (1981) Biochemistry 20, 784–792.
[59] Ogasahara, K. and Yutani, K. (1994) J. Mol. Biol. 236, 1227–1240.
[60] Tanford, C. (1970) Adv. Protein Chem. 24, 1–95.
[61] Johnson, C.M. and Fersht, A.R. (1995) Biochemistry 34, 6795–6804.
[62] Scholtz, J.M., Barrick, D., York, E.J., Stewart, J.M. and Baldwin, R.L. (1995) Proc. Natl. Acad. Sci. USA 92, 185–189.
[63] Kuwajima, K., Nitta, K., Yoneyama, M. and Sugai, S. (1976) J. Mol. Biol. 106, 359–373.
[64] Dolgikh, D.A., Gilmanshin, R.I., Brazhnikov, E.V., Bychkova, V.E., Semisotnov, G.V., Venyanimov, S.Yu. and Ptitsyn, O.B. (1981) FEBS Lett. 136, 311–315.
[65] Christensen, H. and Pain, R.H. (1991) Eur. J. Biophys. 19, 221–229.

[66] Christensen, H. and Pain, R.H. (1994) in Mechanisms of Protein Folding (Pain, R.H., ed.), pp. 55–79, Oxford University Press, Oxford.
[67] Dobson, C.M. (1992) Curr. Opin. Struct. Biol. 2, 6–12.
[68] Ptitsyn, O.B. (1995) Curr. Opin. Struct. Biol. 5, 74–78.
[69] Jennings, P.A. and Wright, P.E. (1993) Science 262, 892–896.
[70] Martin, J., Langer, T., Boteva, R., Schramel, A., Horwich, A.L. and Hartl, F.-U. (1991) Nature 352, 36–42.
[71] Wu, L.C., Peng, Z.-y. and Kim, P.S. (1995) Nature Struct. Biol. 2, 281–286.
[72] Wu, L.C., Schulman, B.A., Peng, Z.-y. and Kim, P.S. (1996) Biochemistry 35, 859–863.
[73] Hendrix, T.M., Griko, Y. and Privalov, P. (1996) Protein Sci. 5, 923–931
[74] Hornby, D.P., Whitmarsh, A., Pinabarsi, H. Kelly, S.M., Price, N.C., Shore, P.D., Baldwin, G.S. and Waltho, J. (1994) FEBS Lett. 355, 57–60.
[75] Schulz, G.E. (1979) in Molecular Mechanisms of Biological Recognition (Balaban, M., ed.), pp. 79–94, Elsevier, Amsterdam.
[76] Loh, S.N., Kay, M.S. and Baldwin, R.L. (1995) Proc. Natl. Acad. Sci. USA 92, 5446–5450.
[77] Nishii, I., Kataoka, M. and Goto, Y. (1995) J. Mol. Biol. 250, 223–238.
[78] Missiakis, D., Betton, J.-M., Minard, P. and Yon, J.M. (1990) Biochemistry 29, 8683–8689.
[79] Munro, A.W., Lindsay, J.G., Coggins, J.R., Kelly, S.M. and Price, N.C. (1996) Biochim. Biophys. Acta 1296, 127–137.
[80] Gahn, L.G. and Roskoski, R., Jr. (1995) Biochemistry 34, 252–256.
[81] Levinthal, C. (1968) J. Chim. Phys. 65, 44–45.
[82] Dill, K.A. (1990) Biochemistry 29, 7133–7155.
[83] Price, N.C. (1994) in Mechanisms of Protein Folding (Pain, R.H., ed.), pp. 160–193, Oxford University Press, Oxford.
[84] Kuwajima, K., Yamaya, H., Miwa, S., Sugai, S. and Nagamura, T. (1987) FEBS Lett. 221, 115–118.
[85] Gilmanshin, R.I. and Ptitsyn, O.B. (1987) FEBS Lett. 223, 327–329.
[86] Kalnin, N.N. and Kuwajima, K. (1995) Proteins Struct. Funct. Genet. 23, 163–176.
[87] Jacob, M.D. and Fox, R.O. (1994) Proc. Natl. Acad. Sci. USA 91, 449–453.
[88] Goldberg, M.E., Semisotnov, G.V., Friguet, B., Kuwajima, K., Ptitsyn, O.B. and Sugai, S. (1990) FEBS Lett. 263, 51–56.
[89] Mann, C.J. and Matthews, C.R. (1993) Biochemistry 32, 5282–5290.
[90] Engelhard, M. and Evans, P.A. (1995) Protein Sci. 4, 1553–1562
[91] Radford, S.E., Dobson, C.M. and Evans, P.A. (1992) Nature 358, 302–307.
[92] Dobson, C.M., Evans, P.A. and Radford, S.M. (1994) Trends Biochem. Sci. 19, 31–37.
[93] Neubert, L.A. and Carmack, M. (1974) J. Am. Chem. Soc. 96, 943–945.
[94] Kishore, R., Raghothama, S. and Balaram, P. (1988) Biochemistry 27, 2462–2471.
[95] Guijarro, J.I., Jackson, M., Chaffotte, A.F., Delepierre. M., Mantsch, H.H. and Goldberg, M.E. (1995) Biochemistry 34, 2998–3008.
[96] Elöve, G.A., Chaffotte, A.F., Roder, H. and Goldberg, M.E. (1992) Biochemistry 31, 6876–6883.
[97] Varley, P., Gronenborn, A.M., Christensen, H., Wingfield, P.T., Pain, R.H. and Clore, G.M. (1993) Science 260, 1110–1113.
[98] Chaffotte, A.F., Cadieux, C., Guillou, Y. and Goldberg, M.E. (1992) Biochemistry 32, 4303–4308.
[99] Agashe, V.R., Shastry, M.C.R. and Udgaonkar, J.B. (1995) Nature 377, 754–757.
[100] Dyson, H.J. and Wright, P.E. (1991) Annu. Rev. Biophys. Biophys. Chem. 20, 519–538.
[101] Brown, J.E. and Klee, W.A. (1971) Biochemistry 10, 470–476.
[102] Muga, A., Surewicz, W.K., Wong, P.T.T., Mantsch, H.H., Singh, V.K., Shinohara, T. (1990) Biochemistry 29, 2925–2930.
[103] Oas, T.G. and Kim, P.S. (1988) Nature 336, 42–48.
[104] Fezoui, Y., Weaver, D.L. and Osterhout, J.J. (1994) Proc. Natl. Acad. Sci. USA 91, 3675–3679.
[105] Minor, D.L., Jr. and Kim, P.S. (1996) Nature 380, 730–734.
[106] Hamada, D., Kuroda, Y., Tanaka, T. and Goto, Y. (1995) J. Mol. Biol. 254, 737–746.
[107] Zitzewitz, J.A., Bilsel, O., Luo, J., Jones, B.E. and Matthews, C.R. (1995) Biochemistry 34, 12812–12819.
[108] Weissman, J.S. (1995) Chem. Biol. 2, 255–260.
[109] Matthews, C.R. (1993) Annu. Rev. Biochem. 62, 653–683.
[110] Schindler, T., Herrler, H., Marahiel, M.A. and Schmid, F.X. (1995) Nature Struct. Biol. 2, 663–673.
[111] Kragelund, B.B., Robinson, C.V., Knudsen, J., Dobson, C.M. and Poulsen, F.M. (1995) Biochemistry 34, 7217–7224.
[112] Sosnick, T.R., Mayne, L. and Englander, S.W. (1996) Proteins: Struct. Funct. Genet. 24, 413–426.
[113] Fersht, A.R., Itzhaki, L.S., El Masry, N.F., Matthews, J.M. and Otzen, D.E. (1994) Proc. Natl. Acad. Sci. USA 91, 10426–10429.
[114] Schmid, F.X., Mayr, L.M., Mucke, M. and Schonbrunner, E.R. (1993) Adv. Protein Chem. 44, 25–66.
[115] Goldenberg, D.P. (1992) Trends Biochem. Sci. 17, 257–261.
[116] Freedman, R.B., Hirst, T.R. and Tuite, M.F. (1994) Trends Biochem. Sci. 19, 331–336.
[117] Hayer-Hartl, M.K. and Hartl, F.-U. (1993) FEBS Lett. 320, 83–84.
[118] Kraulis, J.P. (1991) J. Appl. Crystallogr. 24, 946–950.